View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Bioinformatics

1042-Pos Board B797
Functional Based Analysis and Visualization of Gene Expression Data from Hepatocytes Grown on Different Substrates
Shripad Joshi, Ahmad Al-Zoubi, Aravind Rammohan, Ronald Faris.
Corning Inc., Corning, NY, USA.
The ability to grow a fully functional Hepatocyte *in vitro* would be a prodigious breakthrough for modern day drug testing and regenerative medicine. However there are many factors affecting Hepatocyte functionality *in vitro* that need to be studied before this is possible, like the effect of the underlying substrate that the Hepatocyte is grown on. To better understand this interaction, gene expression profiles of Hepatocytes were measured on four different substrates: Corning(r) CellBIND(r), Collagen, Corning Matrigel(r) and Locust Bean Gum (S906). Here, we present data analysis methods that enable data reduction, identification of differentially expressed genes, and grouping of genes with similar intensities by implementing data clustering. The optimum number of clusters in a given data set is estimated using the gap statistic method. We further employ Cytoscape to help visualize the data and incorporate GeneMania to help identify potentially interesting patterns across substrates. One of the interesting findings from our analysis suggests up-regulation of genes involved in vasculature development on all four substrates. Such findings can be used to design new experiments that can provide further insight on the interactions between hepatocytes and their substrates and help understand the mechanism behind responses of hepatocytes on these substrates.

1043-Pos Board B798
Rheostats and Toggle Switches for Modifying Protein Function
Sarah Meinhardt, Michael W. Manley, Jr, Daniel J. Parente,
Liskin Swint-Kruse.
KU Medical Center, Kansas City, KS, USA.
The millions of protein sequences generated by genomics are expected to transform protein engineering and personalized medicine. To that end, multiple sequence alignments (MSA) are commonly used to identify positions that are conserved during evolution, thereby inferring broad functional properties of new sequences or catastrophic mutations. In addition, nonconserved amino acids can contribute to protein function, and a variety of algorithms have been developed for to identify important nonconserved positions. We previously showed that these algorithms generate true positive and false negative predictions and proposed that algorithm development might be improved by increased experimental knowledge of nonconserved positions.
To that end, we carried out high-throughput, parallel mutagenesis and functional characterization of 12 nonconserved positions in 15 synthetic LacI/GalR homologs, which includes paralogs, orthologs, and isorepressors. Unlike the ''toggle switch'' behaviors of conserved positions, substitutions at nonconserved positions could be rank-ordered to show a ''rheostatic'' effect on function. Amino acid preferences of rheostat positions were highly context-dependent, showed few correlations with physico-chemical similarities, and were not predictable from their occurrence in natural LacI/GalR sequences. Comparisons to bioinformatics predictions suggest that conserved and strongly co-evolving positions act as functional toggles, whereas other important, nonconserved positions serve as rheostats for modifying protein function. These different outcomes should be considered when engineering protein modifications or predicting the impact of protein polymorphisms.

1044-Pos Board B799
Enhancing B-Cell Epitope Predictions by Integrating Protein Sequence and Structural Bioinformatics
Steven J. Darnell, Martin Riese, Erik G. Edlunde, Frederick R. Blattner.
DNASTAR, Inc., Madison, WI, USA.
Monoclonal antibodies specific to unique antigens are invaluable biotechnological tools for diagnosing and treating human diseases. Their laborious production process creates a clear pharmaceutical and research need for tools that guide experiments toward discovering the most promising antigenic targets. A critical barrier toward accurately predicting linear B-cell epitopes (the part of an antigen recognized by an antibody) is the inability to deduce conformational characteristics in the absence of a known 3D protein structure. Most existing prediction methods choose to focus on sequence-derived features or a single structural property, which fail to properly characterize the biophysical mechanisms that mediate protein-protein interactions. We present a knowledge-based model based on sequence, structure, and protein dynamics properties that improves the representation of antibody-antigen interactions and predicts epitopes more accurately than three other leading algorithms. Our current model accurately classifies epitope residues from ''non-epitope'' residues using a machine learning technique trained to recognize the optimal combination from multiple protein features, including secondary structure, local flexibility and rigidity, solvent accessibility, hydrogen bonding potential, antigenicity, and hydrophobicity. As demonstrated by cross-validation analysis and validation with an independent data set, our model displays improved overall predictive accuracy than COBEpro, ElliPro, or Epitopia. In order to enhance the epitope prediction process, we provide a software pipeline that combines our B-cell epitope prediction model with 3D structure prediction (based on protein threading and ab initio modeling techniques) and molecular visualization. This combination enables more accurate epitope predictions for proteins lacking an experimental structure. Our approach aims to elevate the technical capability of a broad range of research scientists in order to facilitate the engineering of more potent monoclonal antibodies.

1045-Pos Board B800
A Search for the Common Words within the Voluminous Phage Vocabulary
Gita Mahmoudabadi.
Caltech, pasadena, CA, USA.
4^{900} is a number that surpasses the estimated number of protons in our observable universe by roughly 462 orders of magnitude. Yet, this number corresponds to the number of sequences that can be encoded using just 900 DNA base pairs, the average gene length! Astonishingly, despite the incomprehensible sequence diversity that the DNA alphabet allows for, there exist sequences that are highly conserved across all domains of life. Such sequences have been used as ''universal'' markers for both identification and evolutionary classification of organisms except for phages (and viruses in general), which are the most abundant of biological entities on our planet, with an estimated 10^{31} such viruses populating the world's ecosystems. In the absence of universal phage markers, we aimed to search for the most ubiquitous phage sequences within a given microbial environment, using the human mouth as an important and intriguing case study. Upon identifying ubiquitous phage markers using a bioinformatic search through oral metagenomic databases, we designed primers to capture them experimentally. To our surprise, using DNA amplification and sequencing, we have verified the presence of these phage markers in all human patients we have sampled so far. Using these markers, we have begun to study phage sequence diversity and evolutionary relationships across different patients, various other animals and natural environments in which these markers are also present. We further aim to use these markers in microsocopy studies where we can directly visualize the networks of phages and their particular hosts in their intact microbial communities.

1046-Pos Board B801
Homcos : A Server to Search and Model 3D Structures of Protein-Protein and Compound-Protein Complexes
Takeshi Kawabata, Haruki Nakamura, Akira Kinjo.
Institute for Protein Research, Osaka University, Osaka, Japan.
3D complex structures of proteins and other molecules provide a clue for mechanism of interaction. However, few servers for searching complex structures are available. We are developing an updated server HOMCOS (HOmology Modeling of Complex Structure) with more services than the previous version. It separates component molecules of PDB files of complexes, such as proteins, nucleic acids, small chemical compounds, stores their binding relationships. It searches these molecules by *BLAST* and our chemical structure comparison program *dkcombu*. Based on found similar complexes, simple template-based models of the complex can be generated by replacing the template with the query amino acid sequence. The service 'Modeling Homo Protein Multimer' searches homologous protein complexes with one query protein sequence, whereas the service 'Modeling Hetero Protein Multimer' searches with two sequences. The service 'Modeling Compound-Protein Complexes' searches with one sequence and one chemical structure. For modeling a bound conformation of a target compound, our flexible overlay program *fkcombu* is performed to superimpose the target compound onto the template compound bound to a protein. The program transforms a pose and torsion angles of the target molecule to superimpose atom pairs of the target and template molecule determined by 2D-MCS. The service 'Searching Contacting Molecules for Query Protein' searches contacting molecules with homologous proteins to a given query sequence, and provide putative interacting molecules with putative binding sites. It is useful to estimate mutational effects on molecular interactions. The service 'Searching Contacting Proteins for Query Compound' searching contacting proteins with similar compounds to a given query chemical compound. It may be useful to predict unintended interacting proteins, which cause side effects of the compound. We hope this server may contribute to the platform for drug discoveries and life sciences.